Simultaneous genetic analysis of longitudinal means and covariance structure in the simplex model using twin data

Dolan, C.V.; Molenaar, P.C.M.; Boomsma, D.I.

published in
Behavior Genetics
1991

DOI (link to publisher)
10.1007/BF01067666

document version
Publisher's PDF, also known as Version of record

Behavior Genetics, Vol. 21, No. 1, 1991

Simultaneous Genetic Analysis of Longitudinal Means and Covariance Structure in the Simplex Model Using Twin Data

Conor V. Dolan,[1] Peter C. M. Molenaar,[1] and Dorret I. Boomsma[2]

Received 30 Sept. 1988—Final 10 Sept. 1990

A longitudinal model based on the simplex model is presented to analyze simultaneously means and covariance structure using univariate longitudinal twin data. The objective of the model is to decompose the mean trend into components which can be attributed to those genetic and environmental factors which give rise to phenotypic individual differences and a component of unknown constitution which does not involve individual differences. Illustrations are given using simulated data and repeatedly measured weight obtained in a sample of 82 female twin pairs on six occasions.

KEY WORDS: repeated measures; genetic and environmental covariance structure; mean trend; longitudinal twin data; genetic simplex mode; LISREL.

INTRODUCTION

Longitudinal data can be examined from two distinct perspectives: the structure and stability of individual differences and the form and continuity of the average growth curve (Wohlwill, 1977). McCall (1981) has stressed the importance of considering both the causes of individual differences and the trend in the species-specific developmental function in studying development. Moreover, McCall points out that the trend in the mean and the individual differences may well be related (see also Thomas, 1980).

[1] University of Amsterdam, Department of Psychology, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands.
[2] Free University, Department of Experimental Psychology, De Boelelaan 1115, 1081 HV Amsterdam, The Netherlands.

0001-8244/91/0100-0049$06.50/0

In behavior genetics, human development is viewed from the perspective of individual differences (Plomin, 1986). This perspective, with its emphasis on the association between, on the one hand, phenotypic differences and, on the other, genetic and environmental differences, does not address the species-specific developmental function. The emphasis on second-order moments in human behavior genetics springs from the limitations inherent in studying behavior in a genetically heterogeneous population where no differential predictions can be made regarding first-degree moments of phenotypes measured within current, or between successive, generations (Mather and Jinks, 1977). Specifically, without the availability of true-breeding lines and the descendants of crosses made between them, it is impossible to estimate genetic and environmental influences on the basis of first-order phenotypic moments.

The situation in human behavior genetics changes, however, if one considers genetic and environmental influences underlying individual differences in multivariate phenotypic measurements. After having determined these genetic and environmental influences by means of a standard genetic analysis of covariance structure (e.g., Martin and Eaves, 1977), it is possible to test the hypothesis that the same influences can also account for the multivariate mean profile, i.e., the differences in univariate phenotypic means making up the multivariate mean vector. In a previous paper, this hypothesis was examined in the context of the static common factor model using twin data (Dolan *et al.*, 1989). In comparison with the approach set out in that paper, however, a genetic analysis of longitudinal means and covariance structure is attended by special difficulties. The issues at stake are discussed in the next sections, followed by illustrative applications to simulated and real data.

The approach to the simultaneous analysis of means and covariance structure used in this paper is based on the work of Sörbom. Sörbom has shown how the difference in observed means between groups may be decomposed into differences in latent variable means in the static factor model (Sörbom, 1974) and in a longitudinal model (Sörbom, 1976; see also Hanna and Lei, 1985). Sörbom's approach can be applied to data obtained from relatives to test the role of biometric latent variable in a difference in means between male and female twins or parents and offspring (Dolan *et al.*, 1990). In the present paper, however, we apply Sörbom's approach to means observed in a single group where the issue is not the decomposition of the difference in means between groups, but the decomposition of the changes in mean trend in a single group of twins.

THE GENETIC SIMPLEX

Consider a longitudinal design involving T consecutive measurement occasions, where a univariate phenotype is repeatedly measured at each occasion $t=1,\ldots T$. We restrict attention to univariate phenotypic time series $\mathbf{P}(t)$, because these present the most intricate difficulties in a simultaneous genetic analysis of longitudinal means and covariance structure. The first step in such an analysis, then, consists of the determination of the genetic and environmental influences by a standard analysis of covariance structure. Given that the phenotypic series constitute realisations of the same developmental process (i.e., constitute an ensemble of time series), the same holds for the underlying genetic and environmental structure. Accordingly, our analysis of covariance structure is based on the genetic (Markovian) simplex model involving nonstationary first-order autoregressive time-series models of these latent influences (Eaves *et al.*, 1986; Boomsma and Molenaar, 1987).

$$\mathbf{P}(t)=\mathbf{G}(t)+\mathbf{E}(t)+\epsilon(t), \qquad t=1,\ldots T, \tag{1}$$

where $\mathbf{G}(t)$ and $\mathbf{E}(t)$ denote genetic and within-families environmental series, respectively, $\epsilon(t)$ represents occasion-specific influences including measurement error, and $E[\mathbf{P}(t)]=E[\epsilon(t)]=0$.

$$\mathbf{G}(t)=\beta(g)_t\mathbf{G}(t-1)+\zeta_g(t), \tag{2a}$$

$$\mathbf{E}(t)=\beta(e)_t\mathbf{E}(t-1)+\zeta_e(t), \tag{2b}$$

where $\beta(g)_t$ and $\beta(e)_t$ are autoregressive (transmission) coefficients, while $\zeta g(t)$ and $\zeta e(t)$ denote random zero-mean innovations which are uncorrelated with $\mathbf{G}(t-1)$ and $\mathbf{E}(t-1)$.

Details concerning the analysis of covariance structure by means of the genetic (Markovian) simplex model are given by Boomsma and Molenaar (1987). Notice that this model can include other, e.g., between-families environmental influences. Also, Eqs. (2) can be placed by other higher-order autoregressive-moving averages (Box and Jenkins, 1970). One thus arrives at a large class of genetic time-series models in which model selection for a given set of empirical covariance matrices proceeds according to the usual criteria such as the likelihood-ratio test and Akaike's (1987) information criterion. For our present purpose, however, it suffices to restrict attention to the simplest model in this class as defined by Eqs. (1) and (2).

THE GENETIC SIMPLEX WITH STRUCTURED MEANS

The present attempt to arrive at a decomposition of the longitudinal mean and covariance structure into genetic and environmental compo-

nents has to be distinguished from standard biometrical analyses of first-order phenotypic moments. In the latter analyses a genetic model is determined from the mean deviation of the true-breeding lines with respect to an origin reflecting the general genetic and environmental circumstances of the observation (cf. Mather and Jinks, 1977, p. 32). Accordingly, this origin plays only a subsidiary role in the analyses concerned, while the identification of genetic and environmental components of the group deviations from this origin is based on knowledge of the genetic constitution of the subjects.

In contrast, the identification of genetic and environmental components of repeatedly observed phenotypic means in a human sample cannot be established in this manner. A decomposition of the phenotypic mean trend can, however, be accomplished when the same genetic and environmental influences are hypothesized to underlie both the phenotypic individual differences as well as the developmental mean curve. More specifically, it is hypothesized that the nonstationary first-order autoregressive models given by Eqs. (2a) and (2b) can account for time-dependent changes in both the longitudinal mean and the covariance structure.

The first step in making this hypothesis testable is to extend Eqs. (2a) and (2b) in such a way that they can accommodate genetic and environmental mean trends. A general way to accomplish this is by letting the means of genetic and environmental innovations become time-varying: $\zeta_g(t) \sim N\{E[\zeta_g(t)], \Psi^2{}_g(t)\}$, and $\zeta_e(t) \sim N\{E[\zeta_e(t)], \Psi^2{}_e(t)\}$, where $N\{\mu, \sigma^2\}$ denotes the normal distribution with mean μ and variance σ^2. Consequently,

$$E[\mathbf{G}(t)] = \beta(g)_t E[\mathbf{G}(t-1)] + E[\zeta_g)t], \tag{2c}$$

$$E[\mathbf{E}(t)] = \beta(e)_t E[\mathbf{E}(t-1)] + E(\zeta_g(t)]. \tag{2d}$$

In Eqs. (2a) and (2c), both the mean and the individual random realizations of the genetic innovation process at time point t are transmitted to the next time point, $t+1$, according to the same autoregressive model (similar remarks apply to the environmental innovations).

It turns out, however, that Eqs. (2c) and (2d) constitute underidentified models of the genetic and environmental mean trends, which will therefore always fit the data. In order to obtain genuinely testable models, one has to narrow down the original hypothesis by imposing further constraints on the time variation of genetic and environmental means innovations. One way in which this can be accomplished is by positing a linear relationship between the mean and the standard deviation of the innovations at each time t: $\zeta_g(t) \sim N\{\Psi_g(t)\Delta(g), \Psi^2{}_g(t)\}$, and $\zeta_e(t) \sim$

$N\{\Psi_e(t)\Delta(e),\ \Psi^2{}_e(t)\}$,where the terms $\Delta(g)$ and $\Delta(e)$ represent time-invariant coefficients of proportionality. Accordingly, Eqs. (2c) and (2d) are replaced by

$$E[\mathbf{G}(t)] = \beta(g)_t E[\mathbf{G}(t-1)] + \psi_g(t)\Delta(g), \tag{2e}$$

$$E[\mathbf{E}(t)] = \beta(e)_t E[\mathbf{E}(t-1)] + \psi_e(t)\Delta(e). \tag{2f}$$

The restricted model given by Eqs. (2e) and (2f) accounts for time-dependent changes of the latent genetic and environmental developmental curves. The phenotypic mean trend consists of a linear combination of these curves superimposed on a constant level ν:

$$E[\mathbf{P}(t)] = \nu + E[\mathbf{G}(t)] + E[\mathbf{E}(t)]. \tag{3}$$

The decomposition of the phenotypic means presented can thus be viewed as a decomposition of the variation of the means about the overall level ν. The parameter ν expresses the time-invariant mean effect of the genetic and environmental influences which do not give rise to time-dependent mean differences in $\mathbf{P}(t)$. Also, the parameter ν allows $\mathbf{P}(t)$ to be measured on an interval scale where the origin of measurement may be changed arbitrarily. Any such change in the origin of the measurements will be absorbed by this parameter without altering the contribution of $\mathbf{E}(t)$ and $\mathbf{G}(t)$ to the mean trend. So the parameter ν consists of two indistinguishable components: (a) the constant effects of genetic and environmental influences which do not contribute to the individual differences and which combine in an unspecified manner and (b) the contribution of an arbitrary measurement origin.

In an nutshell, the genetic simplex model with structured means defined by Eqs. (1), (2a), (2b), (2e), (2f), and (3) is based on the composite hypothesis that (a) the same genetic and environmental processes underlying the longitudinal covariance structure of the phenotypic individual differences also account for the time-dependent changes in the longitudinal mean curve, where (b) the means and variances of these underlying processes are linearly related (cf. McCall, 1981; Thomas, 1980), while (c) genetic and environmental influences specific to the longitudinal mean have a time-invariant effect. Notice that part b of the composite hypothesis can be replaced by a suitable alternative and becomes superfluous in the case of multivariate phenotypic time-series data are available (cf. Dolan *et al.*, 1989).

Furthermore, the requirement of a time-invariant overall level (part a) can be relaxed as the length of the observed time series increases. For instance, given the latent time series presented above and given $T = 6$ measurement occasions, the overall level itself can become time varying.

It would be possible to introduce one level at occasions 1 to 3 and another at occasions 4 to 6 (see below). Knowledge concerning the variable observed or concerning the lengths of the interoccasion intervals may be helpful in the choice of the number and the location (in the series) of the overall level parameters.

APPLICATION TO SIMULATED AND REAL TWIN DATA

The model discussed above can be formulated as a LISREL model (Jöreskog, 1977) for the analysis and monozygotic (MZ) and dizygotic (DZ) between and within mean squares and cross-product (MSCP) matrices. When the mean vector is unconstrained, i.e., in the standard analysis of MSCP structure, we follow the approach given by Boomsma and Molenaar (1987) to the formulation of the genetic simplex model as a LISREL model. The introduction of the mean structure requires the calculation of so-called augmented moment matrices. The definition of augmented MSCP matrices and a detailed specification of the genetic simplex with and without structured means as LISREL models are given in the Appendix.

First, the model was explored using simulated data. To this end 100 MZ and 100 DZ twin pairs were generated for $T = 6$ occasions using the IMSL routine FTGEN (IMSL, 1979). The phenotypic series consisted of an additive genetic series and a specific (unshared) environmental series which were uncorrelated. Both series contributed equally to the phenotypic variance, which was chosen to equal 200 at each occasion (heritability was therefore constant through out the series at .5). The data were analyzed in LISREL VI (Jöreskog and Söbom, 1984) using maximum-likelihood (ML) estimation (this method of estimation was used throughout). The true and recovered parameter estimates are given in Table I. It can be seen that the estimates are close to the true values and that the overall χ^2 is good [$\chi^2(62) = 61.9, p. = 0.52$]. The parameter estimates were all significant judging by their standard errors.

Subsequently the phenotypic means were introduced by augmenting the MSCP between-twin pair matrices in the manner explained in the Appendix. In the analysis the parameters $\Delta(e)$ and $\Delta(g)$ were estimated along with the autoregressive coefficients $\beta(e)_t$ and $\beta(e)_t$. The other parameters, i.e., the innovation variances and their standard deviations, were taken from the previous analysis and introduced as fixed parameters. We do, however, subtract 12 degrees of freedom for these fixed parameters, as they are fixed at estimated values taken from the previous analysis. The parameters $\Psi_e(t)$ and $\Psi_g(t)$, the standard deviations of the

Table I. Results of the Analysis of Simulated data[a]

Parameter	True	Cov, no means	Cov with means
$\beta(g)_1$	.5	.429	.381
$\beta(g)_2$	.6	.591	.631
$\beta(g)_3$	.7	.711	.675
$\beta(g)_4$	.8	.713	.737
$\beta(g)_5$	.9	.985	1.00
$\beta(e)_1$	.9	.922	.934
$\beta(e)_2$	.8	.891	.867
$\beta(e)_3$	.7	.662	.681
$\beta(e)_4$	.6	.641	.631
$\beta(e)_5$	.5	.530	.517
$\Psi^2_g(1)$	100	88.10	Fixed[b]
$\Psi^2_g(2)$	75	75.01	Fixed
$\Psi^2_g(3)$	64	53.82	Fixed
$\Psi^2_g(4)$	51	43.90	Fixed
$\Psi^2_g(5)$	36	48.39	Fixed
$\Psi^2_g(6)$	19	11.94	Fixed
$\Psi^2_e(1)$	100	97.92	Fixed
$\Psi^2_e(2)$	19	15.96	Fixed
$\Psi^2_e(3)$	36	42.71	Fixed
$\Psi^2_e(4)$	51	59.27	Fixed
$\Psi^2_e(5)$	64	70.80	Fixed
$\Psi^2_e(6)$	75	95.98	Fixed
Δg	5		4.383 (2.12)
Δe	−5		−3.245 (1.87)
ν	100		90.927 (7.39)
Goodness of fit		$\chi^2(62) = 61.9$	$\chi^2(71) = 62.9$
Probability.		.52	.74

[a] Maximum likelihood estimates obtained from LISREL. Standard errors for estimates of Δg, Δe, and ν are given in parentheses. All other standard errors are omitted (all estimates are statistically greater than zero and do not deviate significantly from their true values).

[b] Fixed to values obtained from the analysis of covariance structure without structured means, i.e., to the values under "Cov, no mean."

innovation terms, contribute to the latent mean trend in the manner described in Eqs. (2e) and (2f).

The results of the analyses are shown in Table I. It can be seen that the overall goodness of fit is acceptable, with $\chi^2(71)$ of 62.9 (see the Appendix for the calculation of the degrees of freedom in this and subsequent analyses of augmented MSCP matrices). The autoregressive coefficients are similar to those obtained in the analysis with unconstrained means. The parameters $\Delta(g)$ and $\Delta(e)$ are estimated as 4.38(SE = 2.12) and −3.24(SE = 1.87). The overall level was estimated as 90.92(SE = 7.39), which is close to the true value of 100.

Repeating the analysis with the parameter $\Delta(e)$ fixed at zero yielded a χ^2 of 104.23 on 72 df. The difference between the goodness of fit with and without the parameter $\Delta(e)$ equals $\chi^2 = 41.3$ on one degree of freedom ($p < 0.01$). When $\Delta(g)$ was fixed at zero, we obtained a χ^2 of 296.1 on 72 df. Clearly the model correctly detects both the genetic and the environmental contributions to the mean trend.

Finally, the genetic simplex model with structured means is applied to data consisting of repeatedly measured weight. The measures are taken from a larger data set obtained by Dr. S. Fischbein (see Fischbein, 1977). The data used here comprise weight obtained on six equidistant occasions in a sample of 51 DZ and 32 MZ female twins. On the first occasion the average age was 11.5 years (SD = .39) and the interoccasion interval equaled 6 months. The marginal distributions of the data did not show any departures from normality. The MSCP matrices are shown in Table II along with the mean trend.

First, the covariance structure was anlayzed. The genetic simplex model with an additive genetic and an unshared environmental series was found to fit reasonably well, $\chi^2(62) = 74.87$ ($p = .13$). Table III contains the estimated parameters with their associated standard errors. These results show that the variance is due mainly to the additive genetic series but that the unshared environmental series makes a significant, but small contribution. The heritabilities defined as the ratio of the genetic variance to the total variance at the successive occasions are fairly stable: .87, .89, .87, .89, .91, and .89. The innovation variances of both latent series are small, implying a large degree of stability of interindividual differences. The genetic correlations between the successive occasion equal about .97, .96, .97, .97, and .97. The environmental correlations are a little lower but still considerable: .94, .88, .87, .89, and .88.

Having established the genetic simplex as an adequate description of the covariance structure, the structured means were introduced. The parameters relating to the latent variances were fixed at the values obtained from the analysis of covariance structure. So, as in the simulation above, the estimated parameters were the autoregressive coefficients, $\beta(e)_t$ and $\beta(e)_t$, and the parameters, $\Delta(g)$ and $\Delta(e)$. The parameter estimates are shown in Table III. The goodness of fit was acceptable, $\chi^2(71) = 76.47$ ($p = .31$). The overall level ν equals 28.69(SE = .862) and the parameters $\Delta(g)$ and $\Delta(e)$ are estimated as 1.611(SE = .25) and .023(SE = .59), respectively.

As such this analysis suggests that the mean trend of average weight above the overall level of 28.69 is entirely due to the same additive genetic influences which underlie the individual differences in weight. The decomposition of the mean trend as obtained from this analysis is

Table II. Means and Mean-Squares and Cross-Product Matrices for Repeatedly Measured Weight (Six Occasions, Interoccasion Interval of 6 Months Starting at Age 11.5

Means MZF:	36.458	38.791	41.484	43.772	46.098	48.183
Means DZF:	35.557	37.775	40.475	42.806	45.275	47.525
			Mean squares and cross products			
			MZ between (df = 31)			
70.7365						
75.1826	83.1750					
76.6457	83.9897	89.9740				
78.6866	86.8269	94.3119	102.1058			
80.0995	88.4632	95.9566	104.3920	110.4911		
78.8358	86.1574	92.4451	99.9251	105.8316	106.3144	
			MZ within (df = 32)			
3.8652						
3.5611	3.5822					
3.6847	3.6150	4.4416				
3.1513	3.1034	3.8920	4.2978			
2.6375	2.4725	3.0783	3.4661	3.5020		
2.5620	2.4316	2.9620	3.2450	3.3356	4.0283	
			DZ between (df = 50)			
40.6088						
40.7031	42.5535					
42.5176	43.9560	49.1821				
42.4429	43.9419	48.0293	50.6948			
42.1552	42.8368	47.8998	49.1970	51.9043		
40.3719	41.0477	45.7466	46.8776	49.0886	49.6775	
			DZ within (df = 51)			
8.3767						
8.9236	10.5936					
10.1928	12.1854	16.4312				
9.5913	11.8081	15.4734	16.3325			
10.2900	12.5787	16.1654	17.0316	19.3133		
10.4993	12.7942	16.6940	18.0531	20.3446	23.7042	

illustrated in Fig. 1. The increasing genetic contributions equal 7.70, 9.96, 12.69, 14.96, 17.34, and 19.35.

As an aside, the ability of the model to accommodate arbitrary changes of measurement origin was tested by adding 100 to the phenotypic means. The result was simply that the overall level ν was estimated as 128.69 (SE = .870), while all other parameter estimates and the χ^2 goodness of fit remained the same.

A more advanced method of analyzing this data set with structured means is by estimating all parameters simultaneously instead of using

Table III. Results of the Analysis of Weight Data[a]

Parameter	Cov, no means	Cov with means	Cov with means (NONLIS)	Additional parameters from nonlinear equality constraints ($\Psi = \sqrt{\Psi^2}$)	
$\beta(g)_1$	1.053 (.02)	1.060 (.02)	1.058 (.02)		
$\beta(g)_2$	1.044 (.03)	1.031 (.03)	1.035 (.03)		
$\beta(g)_3$	1.019 (.03)	1.024 (.02)	1.023 (.02)		
$\beta(g)_4$	1.023 (.03)	1.011 (.02)	1.01 (.02)		
$\beta(g)_5$	.974 (.03)	.986 (.02)	.983 (.02)		
$\beta(e)_1$	.915 (.05)	.917 (.05)	.914 (.05)		
$\beta(e)_2$	1.051 (.09)	1.064 (.08)	1.060 (.09)		
$\beta(e)_3$	.828 (.08)	.812 (.07)	.820 (.07)		
$\beta(e)_4$	.838 (.07)	.846 (.07)	.847 (.07)		
$\beta(e)_5$	1.005 (.09)	.989 (.09)	.992 (.09)		
$\Psi^2_g(1)$	22.87 (3.36)	Fixed[b]	22.83 (3.23)	$\Psi_g(1)$	4.77 (.34)
$\Psi^2_g(2)$	1.25 (0.22)	Fixed	1.29 (.21)	$\Psi_g(2)$	1.13 (.10)
$\Psi^2_g(3)$	2.26 (0.44)	Fixed	2.15 (.43)	$\Psi_g(3)$	1.47 (.15)
$\Psi^2_g(4)$	1.48 (0.37)	Fixed	1.49 (.37)	$\Psi_g(4)$	1.22 (.15)
$\Psi^2_g(5)$	1.89 (0.38)	Fixed	1.83 (.34)	$\Psi_g(5)$	1.35 (.13)
$\Psi^2_g(6)$	1.96 (0.41)	Fixed	2.06 (.42)	$\Psi_g(6)$	1.44 (.15)
$\Psi^2_e(1)$	3.35 (0.81	Fixed	3.39 (.75)	$\Psi_e(1)$	1.84 (.20)
$\Psi^2_e(2)$	.30 (0.08)	Fixed	.30 (.07)	$\Psi_e(2)$	.55 (.07)
$\Psi^2_e(3)$	.92 (0.22)	Fixed	.96 (.25)	$\Psi_e(3)$	.98 (.13)
$\Psi^2_e(4)$	.93 (0.22)	Fixed	.92 (.22)	$\Psi_e(4)$	.96 (.11)
$\Psi^2_e(5)$	.69 (0.17)	Fixed	.71 (.17)	$\Psi_e(5)$	.84 (.10)
$\Psi^2_e(6)$	.97 (0.22)	Fixed	.94 (.23)	$\Psi_e(6)$	.97 (.12)
Δg		1.611 (.25)	1.618 (.29)		
Δe		.023 (.59)	.01 (70)		
ν		28.69 (.86)	28.69 (1.04)		
Goodness of fit	$\chi^2(62) = 74.87$	$\chi^2(71) = 76.47$	$\chi^2(71) = 75.60$		
p	.13	.31	.33		

[a] Maximum-likelihood estimates obtained from LISREL and NONLIS (columns 4 and 5). Standard errors are given in parentheses.

[b] Fixed to values obtained from the analysis of covariance structure without structured means, i.e., to the values under "Cov, no mean."

parameter estimates obtained from the analysis of covariance structure. The approach requires the specification of nonlinear equality constraints which relate the latent variances to the parameters that scale the contributions of the parameters $\Delta(g)$ and $\Delta(e)$ to the latent mean genetic and environmental trends [see Eqs. (2e) and (2f)]. The equality constraints, in which these weights are specified to equal the square root of the latent innovation variances, cannot be specified in LISREL. We therefore used our own program (NONLIS) for multigroup structural equation modeling

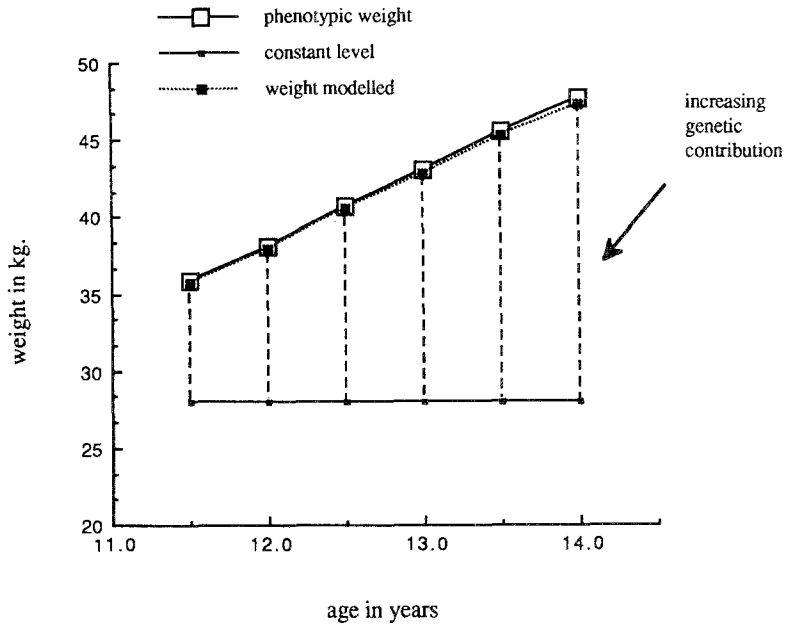

Fig. 1. The decomposition of the mean trend in weight (measured in kilograms) based on the LISREL analysis given in Table III. The phenotypic mean trend shown here is the pooled MZ and DZ mean trends.

to carry out the analysis under the constraints mentioned (again using maximum-likelihood estimation). As can be seen the results (Table III) are almost identical to those obtained from the program LISREL. So although using LISREL is perhaps less elegant from a statistical modeling point of view, it is convenient and yields the same results as obtained from NONLIS.

In the analyses of the weight data carried out so far we have imposed a constant level, ν, throughout the time series. To test the tenability of this aspect we introduced one constant-level parameter (ν_1) or occasions 1 to 3 and a second-level parameter (ν_2) for occasions 4 to 6. The estimates (obtained from NONLIS) equaled $\nu_1 = 28.72$(SE = 1.07) and $\nu_2 = 28.78$(SE = 1.30), while the overall goodness of fit was $\chi^2(70) = 75.59$ ($p = .30$). The goodness of fit for the NONLIS analysis with a single constant level equaled $\chi^2(71) = 75.60$, so we have no reason to reject the hypothesis of a single constant level throughout the time series.

DISCUSSION

A model related to the one presented here is given by McArdle (1986). McArdle's approach differs from the one presented here in a number of important respects. We have based our approach on the (quasi-) Markov simplex model (Jöreskog, 1970; Boomsma and Molenaar, 1987), whereas McArdle employs a level and shape model with second-order biometrical factors. More importantly, our objective is the biometric decomposition of the mean trend. That is, we want to estimate the contribution of genetic and environmental factors to the time-dependent changes in the phenotypic means. McArdle estimates the contribution of the first-order level and shape factors to the phenotypic means (in what seems to be a saturated model in the sense that four parameters are estimated to model four means). However, the biometric factor contributes only to the variance of the level and shape factors, not to their means, so that there is no biometric decompostion of the phenotypic mean trend.

The present approach to the simultaneous biometric analysis of covariance and mean structure is based on the testable assumption that the mean trend can be decomposed into a part that is ascribable to those genetic and environmental factors that contribute to the individual differences and a part that is not. The latter part, which we have referred to above as the constant level, is of an unknown constitution and may include genetic and environmental contributions combining in an unspecified manner. Also, an arbitrary, and therefore meaningless, origin of measurement (allowing measurements to be made on an interval level) is contained in the constant level.

The main problem in the biometric decomposition of the mean trend given herein is the underidentification of the parameters in the autoregressive models for the latent mean trends [see Eqs. (2c) and (2d)]. One is forced, in order to arrive at a testable model, to introduce theoretically motivated constraints. We have chosen to constrain the model by coupling the standard deviations of the innovation variances with the two independent (in the sense that they do not make any contribution to the covariance structure) parameters $\Delta(g)$ and $\Delta(e)$ [see Eqs. (2e) and (2f)]. This approach is based on the assumption that the coupling between means and standard deviations of innovations holds from conception onward. Although there is nothing sacred about this approach, it is simple and intuitively plausible. Also, it has proven to give a good fit to the weight data.

Regarding the results of the analysis of the weight data, these are presented mainly are an illustration of the model. We hope to report a

more elaborate analyses of the Fischbein data both with and without structured mean at a latter date. It is, however, interesting to note the environmental time series as found in the analysis of weight data is consistent with a zero-mean process. This finding is in agreement with the (seemingly arbitrary) assumption given by Falconer (1960) that the contributions of environmental deviations to phenotypes are realizations of a zero-mean Gaussian random variable.

APPENDIX: SPECIFICATION AS LISREL MODELS

The Genetic Simplex Model with Unconstrained Means

Boomsma and Molenaar (1987) use the following submodel in the analysis of MZ and DZ within and between MSCP matrices using LISREL:

$$\Sigma_{ms} = \Lambda(\mathbf{I}-\mathbf{B})^{-1}\,\Psi\,(\mathbf{I}-\mathbf{B}')^{-1}\Lambda', \tag{A1}$$

where Σ_{ms} represents a MSCP matrix. In LISREL four such matrices are specified which differ only in the elements in the matrix Λ. We partition the matrices as follows:

$$\Lambda = \left| \begin{array}{c|c} \Lambda_g & \Lambda_e \end{array} \right|, \tag{A2a}$$

$$B = \left| \begin{array}{c|c} B_g & 0 \\ \hline 0 & B_e \end{array} \right|, \tag{A2b}$$

$$\Psi = \left| \begin{array}{c|c} \Psi_g & 0 \\ \hline 0 & \Psi_e \end{array} \right|. \tag{A2c}$$

The matrix Σ_{ms} can then be shown to the sum of the latent genetic and environmental MSCP matrices:

$$\Sigma_{ms} = \Lambda_g(\mathbf{I}_T-\mathbf{B}_g)^{-1}\Psi_g(\mathbf{I}_T-\mathbf{B}_g')^{-1}\Lambda_g' + \Lambda_e(\mathbf{I}_T-\mathbf{B}_e)^{-1}\Psi_e(\mathbf{I}_T-\mathbf{B}_e')^{-1}\Lambda_e'. \tag{A3}$$

Now, given T occasions, $\Lambda_g(T \times T)$ is a diagonal matrix equaling $\sqrt{w}\,\mathbf{I}_T$. The scalar w represents the genetic weights equaling 2 (between MZ), 0 (within MZ), 1.5 (between DZ), and .5 (within DZ) (Mather and Jinks, 1977). The matrix $\mathbf{I}_T$ is the $(T \times T)$ identity matrix. The matrix $\mathbf{B}_g$ $(T \times T)$ contains on its first lower subdiagonal, i.e., in the positions $\mathbf{B}_g(i, i-1)$, $i = 2,\ldots,T$, the genetic autoregressive coefficients. The matrix $\Psi_g(T \times T)$ is diagonal, containing the variances of the innovation terms $\zeta_g(t)$ [Eq. (2a)].

The matrices of the environmental covariance structure are defined analogously. Here, of course, the matrix $\Lambda_e(T \times T)$ equals $\mathbf{I}_T$.

The Genetic Simplex Model with Structured Means

The introduction of the means into the analysis of MSCP matrices requires the calcuation of the so-called augmented moment matrices, which can be defined, in the present case, as follows:

$$\Sigma_{\text{am}} = \left[\begin{array}{c|c} \Sigma_{\text{ms}} + \mu\mu' & \mu \\ \hline \mu' & 1 \end{array}\right], \tag{A4}$$

where the T-dimensional vector μ is the vector of phenotypic means:

$$\mu = \mu_g + \mu_e + \nu. \tag{A5}$$

That is, the phenotypic means are the sum of the genetic and environmental mean vectors of which the elements are defined in Eqs. (2a) and (2b) and the vector ν, which contains the constant-level parameter. As we have assumed that the constant level is equal throughout the time series, the T-dimensional vector contains identical elements equal to ν [see Eq. (3)]. The matrix Σ_{am} now has the dimensions $(T+1 \times T+1)$.

The following LISREL model is now used to model the augmented MSCP matrices:

$$\Sigma_{\text{am}} = \left[\begin{array}{c|c} \Lambda(\mathbf{I}-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(\mathbf{I}-\mathbf{B}')^{-1}\Lambda' & \Lambda(\mathbf{I}-\mathbf{B})^{-1}\Gamma\Phi\Lambda_x' \\ \hline \Lambda_x\Phi\Gamma'(\mathbf{I}-\mathbf{B}')^{-1}\Lambda' & \Lambda_x\Phi\Lambda_x' \end{array}\right]. \tag{A6}$$

So the T-dimensional phenotypic mean μ vector is modeled as

$$\mu = \Lambda(\mathbf{I}-\mathbf{B})^{-1}\Gamma\Phi\Lambda_x' \ . \tag{A7}$$

The matrices in Eq. (7) can be partitioned as follows:

$$\Lambda = \left[\begin{array}{c|c|c} \Lambda_g & \Lambda_e & \Lambda^* \end{array}\right], \tag{A8a}$$

where $\Lambda_g(T \times T)$ and $\Lambda_e(T \times T)$ have been defined above and the matrix Λ^* $(T \times 4)$ is defined as

$$\Lambda^* = [\mathbf{0}_T\mathbf{0}_T\nu\mathbf{0}_T]. \tag{A8b}$$

Here $\mathbf{0}_T$ represents a T-dimensional zero vector and ν has been defined above as the vector of constant-level parameters.

Next the matrix B is specified as follows:

$$B = \left[\begin{array}{c|c|c} B_g & \mathbf{0} & B_{\text{wg}} \\ \hline \mathbf{0} & B_e & B_{\text{we}} \\ \hline \mathbf{0} & \mathbf{0} & B^* \end{array}\right], \tag{A8c}$$

where the matrices $B_g(T \times T)$ and $B_e(T \times T)$ have been defined above. The matrix B_{wg} $(T \times 4)$ equals

$$B_{wg} = [\mathbf{0}_T \mathbf{W}_g \mathbf{0}_t \mathbf{0}_T]. \tag{A8D}$$

The T-dimensional vector $\mathbf{W}_g$ contains the standard deviations of the genetic innovation terms, $\Psi_g(t)$ [see Eq. (2a)]. The $(T \times 4)$ matrix B_{we} is defined analogously:

$$B_{we} = [\mathbf{0}_T \mathbf{0}_T \mathbf{0}_T \mathbf{W}_e]. \tag{A8e}$$

The matrix $B*$ (4×4) is given as

$$B^* = \left| \begin{array}{cccc} 0 & 0 & 0 & 0 \\ \Delta(g) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \Delta(e) & 0 \end{array} \right|, \tag{A8f}$$

where the terms $\Delta(g)$ and $\Delta(e)$ have been defined above as time-invariant coefficients of proportionality.

The vector Γ equals

$$\Gamma' = [\mathbf{0}_T' \ \mathbf{0}_T' \ \Gamma^{*\prime}], \tag{A8g}$$

with the four-dimensional vector $\Gamma^{*\prime} = [1/\sqrt{w}\ 0\ 1\ 0]$.

The matrix Ψ $(T+4 \times T+4)$ is defined as

$$\Psi = \left| \begin{array}{c|c|c} \Psi_g & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \Psi_e & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right|, \tag{A8h}$$

where $\Psi_g(T \times T)$ and $\Psi_e(T \times T)$ have been defined above.

Finally, the matrices Λ_x (1×1) and Φ (1×1) both equal the scalar 1.0. Working through the matrix algebra of Eq. (A7) yields the expression for the phenotypic means given in Eq. (3) with latent means defined in Eqs. (2a) and (2b).

In the specification of this LISREL model the phenotypic means were introduced only into the between-groups matrices; in the within-groups matrices the vector μ was specified as a zero vector.

$$\Sigma_{am} = \left| \begin{array}{c|c} \Sigma_{ms} + \mu\mu' & \mathbf{0}_T \\ \hline \mathbf{0}_t' & 1 \end{array} \right|. \tag{A9}$$

The within-group matrices cannot be used to introduce the means because there are no genetic within-twin pair effects in the MZ groups so that the matrix Λ_g is zero. Because the within-groups matrices are augmented

as shown in Eq. (A9), the LISREL program evaluates the χ^2 goodness-of-fit index on the basis of a spuriously large number of degrees of freedom (df).

Given T occasions, we have $[T^*(T+1)]/2$ df for each within MSCP matrix and $[T+1^*(T+2)]/2$ df for a between MSCP matrix. From the latter we subtract 1 df for the fixed element in the lower right submatrix in Eq. (A4). So given $T=6$ occasions, the augmented MSCP MZ and DZ matrices yield a total of $2^*[21 + (28-1)] = 96$ df.

ACKNOWLEDGMENTS

We would like to thank Drs. Siv Fischbein, Kay Phillips, and Carol Prescott. Dr. Fischbein kindly made available her twin data. Reviews by Drs. Prescott and Phillips of an earlier version of this paper led to many improvements.

REFERENCES

Akaike, H. (1987). Facator Analysis and AIC. *Psycholmetrika* **53**(3):317–332.

Boomsma, D. I., and Molenaar, P. C. M. (1987). The genetic analysis of repeated measures. I. Simplex models. *Behav. Genet.* **17**:111–123.

Box, G. E. P., and Jenking, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Dolan, C. V., Molenaar, P. C. M., and Boomsma, D. I. (1989). LISREL analysis of twin data with structured means. *Behav. Genet.* **19**:51–62.

Dolan, C. V., Molenaar, P. C. M., and Boomsma, D. I. (1990). Decomposition of group-related differences in multivariate phenotypic means in genetic covariance structure analysis (*submitted for publication*).

Eaves, L. J., Long, J., and Heath, A. C. (1986). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behav. Genet.* **16**:143–162.

Falconer, D. S. (1960). *Introduction to Quantitiative Genetics*, Oliver and Boyd, Edinburgh.

Fischbein, S. (1977). Intra-pair similarity in physical growth of monozygotic and dizygotic twins during puberty. *Ann. Hum. Biol.* **4**:417–430.

Hanna, G., and Lei, H. (1985). A longitudinal analysis using the LISREL-model with structured means. *J. Educ. Stat.* **10**:161–169.

IMSL Inc. (1979). *IMSL Library Reference Manual*, ed. 7, IMSL Inc., Houston, Tex.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *Bri. J. Math. Stat. Psychol.* **26**:121–145.

Jöreskog, K. G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In Krishaiah, P. R. (ed.), *Applications of Statistics*, North-Holland, Amsterdam.

Jöreskog, K. G., and Sörbom, D. (1984). *LISREL VI: Analysis of Linear Structural Relations by the Method of Maximum Likelihood*, National Education Resources. Chicago.

Martin, N. G., and Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity* **38**:79–95.

Mather, K., and Jinks, J. L. (1977). *Introduction to Biometrical Genetics*, Chapman and Hall, London.

McArdle, J. J. (1986). Latent variable growth within behavior genetic models. *Behav. Genet.* **16**:163–201.
McCall, R. B. (1981). Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Dev.* **5**:1–12.
Plomin, R. (1986). *Development, Genetics, and Psychology*, Lawrence Erlbaum, Hillsdale, N.J.
Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *Br. J. Math. Stat. Psychol.* **27**:229–239.
Sörbom, D. (1976). A statistical model for the measurement of change in true scores. In de Gruyter, D. N. M., and van der Kamp, L. J. T. (eds.), *Advances in Psychological and Educational Measurement*, John Wiley & Sons, New York.
Thomas, H. (1980). A theory of Growth. In Kluwe, R. H., and Spada, H. (eds.), *Developmental Models of Thinking*, Academic Press, New York.
Wohlwill, J. F. (1973). *The Study of Behavioral Development*, Academic Press, New York.

Edited by N. G. Martin